DOCUMENT RESUME

ED 353 284 TM 019 002

AUTHOR Marsh, Herbert W.; Byrne, Barbara M.
TITLE Do We See Ourselves as Others Infer: A Comparison of Self-Other Agreement on Multiple Dimensions of Self-Concept from Two Continents.
PUB DATE 13 May 92
NOTE 23p.; For an earlier version of this study, see ED 319 787.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Students; Comparative Testing; Cross Cultural Studies; *Factor Analysis; Foreign Countries; Higher Education; *Inferences; *Peer Evaluation; *Self Concept; Self Evaluation (Individuals); Significant Others; *Test Validity
IDENTIFIERS Australia; Canada; *Self Description Questionnaire III

ABSTRACT

An Australian study and a Canadian study evaluated the ability of significant others to infer multiple dimensions of self-concept of university students accurately. Subjects were 151 Australian and 941 Canadian university students. Four factor analyses (self- and other-responses by Australians and Canadians) all clearly identified the 13 factors that the 136-item Self Description Questionnaire III (SDQIII) is designed to measure. Mean ratings were similar in the two studies, although self-rankings tended to be lower than other ratings (a self-modesty effect). There was substantial self-other agreement on each of the 13 SDQIII factors that was similar for Australians and Canadians and much higher than reported in previous research. Multitrait multimethod data demonstrated convergent and discriminant validity of the SDQIII responses in both studies. Across all analyses, Australian and Canadian results were remarkably similar. Critical features leading to the consistently high self-other agreement were the use of: (1) older students; (2) multiple dimensions of self-concept based on instruments with strong psychometric properties; and (3) significant others who knew the subject very well. Four tables present study data, and 38 references are included. An appendix lists the means and standard deviations among the variables from Australian and Canadian studies. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED353284

Do We See Ourselves As Others Infer: A Comparison of Self-other Agreement On Multiple Dimensions of Self-Concept From Two Continents

Herbert W. Marsh
University of Western Sydney, Macarthur (Australia)

Barbara M. Byrne
University of Ottawa (Canada)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
HERBERT W. MARSH
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

13 May, 1992

Running Head: Self-other Agreement

Requests for further information about this article or the Self Description Questionnaires should be sent to Herbert W. Marsh, Professor of Education, University of Western Sydney, Macarthur, PO Box 555, Campbelltown NSW 2560, Australia.

ABSTRACT

Two studies evaluated the ability of significant others to accurately infer multiple dimensions of self-concept of university students in Australia (n=151) and Canada (n=941). Four factor analyses -- self- and other-responses by Australians and Canadians -- all clearly identified the 13 factors that the Self Description Questionnaire III (SDQIII) is designed to measure. Mean ratings were similar in the two studies, although self-ratings tended to be lower the other-ratings (i.e., a self-modesty effect). There was substantial self-other agreement on each of the 13 SDQIII factor that was similar for Australians and Canadians (mean rs of .568 and .560), and much higher than reported in previous research. MTMM data demonstrated convergent and discriminant validity of the SDQIII responses in both studies. Across all analyses, Australian and Canadian results were remarkably similar. Critical features leading to the consistently high self-other agreement were the use of: older students, multiple dimensions of self-concept based on instruments with strong psychometric properties, and significant others who know the subject very well.

TM 019 002

BEST COPY AVAILABLE

Self-concept ratings by others -- inferred self-concepts -- are used to determine how accurately self-concept can be inferred by external observers, to validate interpretations of responses to self-concept instruments, and to test a variety of theoretical hypotheses such as those derived from the symbolic interactionist perspective. The study of agreement between self-concept ratings and self-concept inferred by others has a long and controversial history (e.g., Baldwin, 1897; Bem & Allen, 1974; Burns, 1979; Cooley, 1902; Duval & Wicklund, 1972; James, 1890; Kinch, 1963; Mead, 1934; Marsh, Barnes & Hocevar, 1985; Marsh & Craven, 1991; Shrauger & Schoeneman, 1979; Wells & Marwell, 1976; Wylie, 1974; 1979). Depending on the aims of the study, it may be relevant to compare the subject's own self-concept ratings with inferred self-concept ratings by others of what subjects would say (i.e., predict the subject's self-concept responses) or ratings by others of what subjects should say (as opposed to what subjects actually do say). The second approach, for example, might be appropriate to determine how accurately subjects view themselves compared to the perceptions of others. Self-concept, however, is based on self-perceptions -- whether accurate or not -- and so the second approach is used most frequently in self-concept research. Consistent with the aims of the present investigation, Marsh, Barnes and Hocevar (also see Wells and Marwell) argued that inferred self-concept ratings by other are the most useful in order to determine how accurately self-concept can be inferred by external observers, and particularly to validate interpretations of responses to self-concept instruments.

Self-other discrepancies in mean ratings by self and by others are sometimes used to test the frequently hypothesized self-favorability bias -- that self-ratings are systematically higher than they "should" be. In such studies the focus is on mean differences between self-responses and responses by others rather than correlations between responses. Such a self-favorability effect on self-concept may represent the influence of selective perceptions and interpretations, memory, the perhaps unrealistic feedback often given particularly to young children, frame of reference effects (i.e., different standards of comparison), or, alternatively, intentional distortions in self-reports that do not accurately reflect true self-perceptions. In a review of this research, Wylie (1979, p. 681) concluded that: "there appears to be a considerable consistency among the methodologically more adequate studies in showing trends toward self-favorability biases regarding evaluative characteristics." Wylie, however, specifically excluded studies based on "private-self-concept responses" (i.e., the subject is instructed to report how he or she privately sees him or herself, whether or not this agrees with external criteria) and only considered studies in which subjects made "social-self-concept responses" (i.e., regardless of his or her own private view of him or herself, the subject is to reporte or she thinks generalized or particular others would characterize him or her) or made self-ratings relative to some objective standard such as school grades. The social-self-concept that she used to evaluate self-favorability biases is not the self-concept considered here, is not the self-concept typically considered in other self-concept research, and is not the self-concept considered by Wylie in other sections of her 1974 and 1979 books. Hence there is need for further research on self-other discrepancies for responses in the form that they are typically used in self-concept research.

Shrauger and Schoeneman (1979) reviewed studies that correlated self-reports with judgments by others, and concluded that "there is no consistent agreement between people's self-perceptions and

how they are actually viewed by others" (p. 549). These conclusions reflect negatively on the construct validity of self-concept responses. However, in the Shrauger and Schoeneman review: the content of the self-reports was quite varied, no attempt was made to determine if some external observers (e.g., teachers, parents, peers) provided more accurate assessments than others, and they did not consider the distinctiveness of different components when multiple characteristics were judged. Furthermore, they did not distinguish between studies that asked external observers to record their own perceptions and those in which observers made inferred self-concept ratings. More recent self-concept research, particularly research reviewed here that is based on responses to the Self Description Questionnaires (SDQ), apparently contradicts Shrauger and Schoeneman's conclusions.

Self-other agreement on multidimensional self-concept ratings

Historically, self-concept research has emphasized a general, global or total self-concept, but more recently there has been growing support for the multidimensionality of self-concept. There is particularly good support for the Shavelson model of self-concept (Shavelson, Hubner & Stanton, 1976; Marsh, Byrne & Shavelson, 1988; Marsh & Shavelson, 1985) and the SDQ instruments based on this model. According to the Shavelson model, self-concept is a multifaceted, hierarchical construct; general self-concept at the apex of the hierarchy is divided into academic and nonacademic self-concepts which are further divided into more content specific components of self. In her review of self-concept models Byrne (1984, p. 449) concluded that: "Although no one model to date has been sufficiently supported empirically so as to lay sole claim to the within-network structure of the construct, many recent studies, in particular those by Marsh and his colleagues, are providing increasingly stronger support for the hierarchical model." Subsequent reviews of research prompted by this model (Marsh, Byrne & Shavelson, 1988; Marsh & Shavelson, 1985; Marsh, 1990a) led to the Marsh/Shavelson revision of the original model in which self-concept was posited to be even more multifaceted and less hierarchically ordered. This research provides further support for the multifaceted structure of self-concept and demonstrates that self-concept cannot be adequately understood if this multidimensionality is ignored.

When multiple dimensions of self-concept are represented by both self-ratings and inferred-ratings, multitrait-multimethod (MTMM) analysis provides an important analytical tool for testing the construct validity of the self-concept facets (see Campbell & Fiske, 1959; Marsh, 1988a; 1989; for more general discussions of MTMM analyses; also see Shavelson, et al., 1976; Wylie, 1979 for MTMM analyses as a method of validating self-concept responses). Convergent validity, the traditional focus of self-other agreement studies, is inferred from substantial correlations between self-ratings and inferred-ratings on matching self-concept traits. Discriminant validity provides a test of the distinctiveness of self-other agreement and of the multidimensionality of the self-concept facets; it is inferred from the lack of correlation between nonmatching traits. MTMM studies using the SDQI for preadolescents and the SDQIII for late-adolescents are briefly reviewed below. (There are three SDQ instrument designed for preadolescents (SDQI), for early and middle adolescents (SDQII), and for late adolescents and young adults (SDQIII) that are summarized in greater detail by Marsh (1990a)).

Preadolescent Studies.

Eight MTMM studies using the SDQI (see Marsh, 1988b) demonstrated significant agreement between multiple self-concepts inferred by primary school teachers and student responses to the SDQI. Across all 8 studies the average of the 56 convergent validities (self-other agreement on matching scales) was .30 (excluding the General self-concept scale that was only considered in two of the studies). Student-teacher agreement was higher in those areas in which relevant behaviors were most observable (math, .37; reading, .37; general school, .33; physical ability, .38; and, peer relations .29), but was lower on Parent Relations (.17) and somewhat surprisingly Physical Appearance (.16). The General self-concept scale was only considered in two studies, but self-other agreement (.15) was very low. Thus, taking General self-concept into account would reduce the average self-other agreement between students and teachers. These studies demonstrate that external observers can infer self-concepts in many areas with at least modest accuracy, thus countering Shrauger and Schoeneman's 1979 claim to the contrary. Whereas support for convergent and discriminant was evident when evaluated by the traditional Campbell-Fiske criteria, the level of self-other agreement was only modest. There are several likely explanations for why self-other agreement is only modest: (a) preadolescents may be more likely than older subjects to base their self-concepts on idiosyncratic criteria that are unobservable or not considered by external observers; (b) teachers may not have an appropriate basis for inferring self-concepts in some areas; and (c) because teachers made ratings of all students in their class, they were only asked to respond to psychometrically weaker single-item scales instead of the multi-item scales completed by students.

More recently, Marsh and Craven (1991) evaluated some of these concerns. They evaluated agreement between self-concept responses by students in grades 3, 4, 5, and 6 and inferred self-concept responses by teachers, mothers and fathers. Unlike the studies reviewed earlier, the inferred self-concept responses were based on responses to the complete SDQI instrument instead of the single-item summary ratings. Despite the somewhat younger ages of children in this study, the results showed consistently better self-other agreement than earlier research (mean $r = .42$ across all 8 SDQI scales and ratings by teachers, mothers and fathers). The pattern of results, however, was similar in that teachers, mothers, and fathers were all more accurate at inferring self-concepts of Maths, Reading, School, Physical Ability, and, to a lesser extent, Peer Relations. Across all three significant others, self-other agreement was lowest for General self and Parent Relations. Surprisingly, there were not large differences in the abilities of the three significant others to infer self-concept -- not even for self-concepts of Parent Relationships, Reading, Maths, and School. The stronger self-other agreement found in this study compared with earlier research underscores the importance of using multi-item scales with good psychometric properties. The Marsh and Craven study also provides a potentially important basis of comparison for research with older subjects to be considered in the present investigation.

Late-adolescent/young adult studies.

A particularly important MTMM study (Marsh, Barnes & Hocevar, 1985) was conducted with SDQIII responses by a small sample 151 of Australian university students. Students were asked to complete the SDQIII and to ask the person in the world who knew them best to complete the SDQIII as if they were the person who had given them the survey. The significant others were typically family

members -- most frequently a parent. Separate factor analyses of both self-ratings and responses by significant others identified the 13 dimensions of self-concept which the SDQIII was designed to measure. Self-other agreement was very high (mean r = .57), and four of the scales had self-other correlations over .75 (Physical ability, Religion/spiritual values, Parent relations, and Mathematics). An application of the traditional Campbell and Fiske (1959) guidelines for evaluating MTMM matrices provided strong support for the convergent and the discriminant validity of the SDQIII responses. As part of the same study, subjects and significant others also responded to single-item summary ratings like those used in many other studies. Whereas there was still support for the convergent and discriminant validity of these single-item responses, self-other agreement was substantially lower than found on the multi-item scales. The self-other agreement based on the multi-item scales in this study was much higher than previously reported. Particularly on the four scales with the highest self-other agreement the results may be the strongest relation between self-reports and an external validity criterion found in personality research.

Marsh and Richards (1990) further examined self-other agreement on SDQIII responses for 280 participants in an Outward Bound program. Participants in this 26-day residential program worked primarily in small groups and observed the other members of their group in many different situations. Each subject chose two group members who knew him or her the best, and these external observers were asked to complete single-item summary ratings as the subject "would" complete them and as the subject "should" complete them. MTMM analyses of agreement between the two external observers indicated modest agreement (mean r = .32) and support for convergent and discriminant validity of their responses. Similarly, agreement between responses by the two external observers and self-responses by the subject (mean r = .37) was moderate. Although the results provided support for convergent and divergent validity, correlations among ratings by external observers to different areas of self-concept were substantially higher than among self-response ratings, suggesting a method/halo effect in external observer responses. External observers were apparently unable to differentiate between "would" and "should" responses; "would" responses by different observers were no more highly correlated than were the "would" and "should" responses by different observers. Although self-other agreement in this study was higher than reported by Shrauger & Schoeneman (1979) it was substantially lower than reported by Marsh, Barnes and Hocevar (1985). The findings should, however, be evaluated in relation to two restrictions; external observers responded to single-item summary scales instead of multi-item scales and had contact, albeit intensive contact, with subjects for only 26 days. Particularly in relation to these limitations, support for the convergent and discriminant validity of the responses was surprisingly good.

Marsh, Barnes and Hocevar (1985) also compared the mean responses by students and by significant others and found a slight tendency for higher responses by significant others. Although not emphasized in the study, these findings may call into question Wylie's 1979 claim for self-favorability biases. This issue was examined in more detail by Marsh and Richards (1990). Self-concept responses in the Marsh and Richards (1990) study were consistently less favorable than those of external observers, suggesting a self-modesty effect instead of a self-favorability bias. This modesty effect was also consistent with the finding that observer "would" ratings were marginally lower than observer

"should" ratings. That is, observers said that subjects "would" give themselves lower self-ratings than they "should". In evaluating this modesty effect it is important to note that self-responses and observer responses were actually made by the same individuals. That is, each participant judged him or herself (self-responses) and also made judgments of two other participants (observer responses). On average then, participants indicated that other participants "would" and "should" give themselves higher ratings than participants gave themselves -- again suggesting a modesty effect in the self-responses. Because the stimulus materials and individuals making the judgments were the same for the self-rating and observer tasks, many influences that might differentially affect self and other ratings (e.g., differences in response biases, the constructs being evaluated, frame of reference effects) are less plausible. In evaluating these results, Marsh and Richards suggested the possibility that modesty effects may be more likely in Australian studies than North American studies where self-favorability effects are typically found (Wylie, 1979).

Self-reports And Ratings In The Assessment Of Personality

Despite the historical importance of self-other agreement to self-concept research, recently there appears to be more systematic research on this issue in other areas of personality research (e.g., Cheek, 1982; Funder, 1987; Kenny & Albright, 1987; Kenrick & Funder, 1988). McCrae and Costa (1988; also see Conley, 1985; Costa & McCrae, 1988) reviewed theoretical implications and empirical findings in studies correlating personality traits derived from self-responses and from responses by others in research particularly relevant to the present investigation. As in self-concept studies emphasized in the present investigation, personality researchers examined self-other agreement on multiple traits to test the construct validity of self-reports and used factor analysis and MTMM analysis as their principal analytic tools. Their research was designed to counter previous claims that correlations between self-reports and ratings by others cannot break the so-called .3 barrier and the widely held beliefs that personality traits based on self-report reflect primarily self-presentation or response biases. Despite these similarities between the personality and the self-concept studies, there are important differences: (a) in personality research, self-responses are designed to provide objective indicators of personality traits and thus are more like what Wylie (1979) referred to as public self-concept rather than the private self-concept that is the focus of most self-concept research; (b) observer ratings in the personality research are designed to provide an accurate appraisal of the subject whether or not they reflect the subject's self-perceptions; that is, they are more like the "should" ratings in the Marsh and Richards (1990) study than the inferred self-concept rating (i.e., "would" ratings) used in self-concept research; (c) even though personality studies often consider multiple traits, the major focus is convergent validity rather than divergent validity. Nevertheless, methodological advances in either area are likely to contribute to other area.

McCrae and Costa (1988) summarized 10 studies in which there was substantial agreement between self-reports and peer ratings. Whereas self-other agreement was consistently less than that reported by Marsh, Barnes and Hocevar (1985), some studies approached this level of agreement by aggregating responses from as many as 10 different observers. When external observers were spouses, however, self-other correlations on five personality domains varied between .5 and .6, though correlations based on subscales within each domain were somewhat lower (Costa & McCrae, 1988).

In the same study, correlations between self-ratings and peer ratings, and between spouse ratings and peer ratings were lower (median r = .42), apparently reflecting the fact that spouses knew the subjects better than peers. The McCrae and Costa study based on spouse ratings and the Marsh, Barnes and Hocevar study based on responses by persons selected by subjects as knowing them the best both suggest that knowledge of the subject is a critical variable. McCrae and Costa (1988) concluded their review of self-other agreement by noting that: "When reliable and valid measures are used, the correlations considerably exceed the .3 barrier; they are better characterized as facing a .6 barrier" (p. 5). However, whereas none of the correlations reported by Costa and McCrae was larger than .6, Marsh, et al. reported self-other correlations in excess of .75 for 4 of 13 self-concept scales.

Summarizing characteristics that seem to enhance self-other agreement, McCrae and Costa (1988) noted the use of: (a) multi-item scales instead of single-item ratings, (b) instruments with superior psychometric properties, (c) factor analytically derived factor scores that maximally distinguish between the multiple dimensions being considered, and (d) responses from observers who know the ratee better in a variety of different contexts. In a proposal similar to that of Shavelson et. al. (1976), they also suggested that agreement may be better on traits that are more closely linked to observable behavior. Hence, characteristics that enhance correlations between self-responses and external observer ratings of personality traits appear to be similar to those that lead to higher correlations between self-concept ratings and inferred self-concept ratings.

The Present Investigation

We hypothesize that when critical features of the Marsh, Barnes and Hocevar (1985) Australian study are replicated in a large sample of North American university students, self-other agreement on multiple dimensions of self-concept will be comparable to that found in the Australian study and higher than found in other self-concept research. This study is important because: (a) whereas SDQIII responses have been used widely in Australian studies, its applicability to North American samples has not been tested; (b) the small sample (N=151) in the Australian study may call into question its generalizability, particularly given that it found self-other agreement so much higher than found in other self-concept studies; and (c) Marsh and Richards (1990), commenting on their finding of a modesty bias (self-ratings lower than ratings by others) that ran counter to the favorability bias typically found in North American studies, suggested that Australians may be more modest in evaluating themselves than North Americans.

Methods

Sample and Procedures.

Australian study. This study is described in detail by Marsh, Barnes and Hocevar (1985) and is summarized here only briefly. The sample consisted of 151 Australian university students from psychology and education classes who volunteered to participate in the study. Students completed the SDQIII and then asked the person in the world who knew them best to complete the companion survey using the same instructions as were subsequently used in the Canadian study. Because students in this study often lived at home, the significant others were typically family members and over half were parents.

Canadian study. Subjects were 941 introductory psychology students in a large Canadian university who volunteered to participate as partial fulfillment of a course requirement. As in the Australian study, students completed the SDQIII and then were asked to chose the person in the world who knew them best to also complete the SDQIII. On the companion survey, the significant others were asked to imagine that they were the person who had given them the survey and to complete the same SDQIII items as if they were that person. Students were explicitly instructed not to discuss the survey with their selected significant other. A pre-addressed envelope was included with the survey that was given to the significant other and they were explicitly instructed to return the survey without discussing their responses with the student. Although the relation between students and the significant other was not obtained, informal queries suggested that about half of the significant others were intimate partners (spouse, live-in partners, or boy friend/girl friend) and that most others were an immediate family member -- typically a parent. A total of 1081 students completed the SDQIII and 941 pairs of instruments completed by the subject and by the significant other were obtained. Analyses described here are based on the 941 pairs of response where both self-responses and responses by the significant other were completed.

The SDQIII Instrument.

The 136-item SDQIII instrument assesses 4 areas of academic self-concept, 8 areas of nonacademic self-concept, and a General Self scale derived form the Rosenberg (1965) self-esteem instrument. On the SDQIII, late adolescents or young adults are asked to respond to simple declarative sentences (e.g., I am good looking; I worry a lot; I have trouble with most academic subjects) with one of 8 responses: Definitely False; False; Mostly False; More False Than True;More True Than Fales; Mostly True; True; Definitely True. Each of the 13 SDQIII scales is inferred on the basis of responses to 10 or 12 items, half of which are negatively worded. Norms based on responses by a total of 2,436 Australian subjects are presented in the test manual (Marsh, 1992) along with a detailed presentation of the theoretical and empirical basis of the instrument. The psychometric characteristics of responses to the SDQIII -- reliability and factor structure -- based on the normative archive are compared with those based on the Australian and Canadian studies considered as part of subsequent analyses. The 13 SDQIII scales are summarized as follows:

(1) Physical Abilities: Student perceptions of their skills and interest in sports, games, and physical activities.

(2) Physical Appearance: Student perceptions of their physical attractiveness, how their appearance compares with others, and how others think they look.

(3) Opposite Sex Relationships: Student perceptions of their popularity with members of the opposite sex, how easily they make friends with members of the opposite sex, and the quality of their interactions with members of the opposite sex.

(4) Same Sex Relationships: Student perceptions of their popularity with members of the same sex, how easily they make friends with members of the same sex, and the quality of their interactions with members of the same sex.

(5) Honesty/Trustworthiness: Student perceptions of their honesty, reliability and trustworthiness.

(6) Parent Relations: Student perceptions of how well they get along with their parents, whether they like their parents, and the quality of their interactions with their parents.

(7) Spiritual Values/Religion: Student self-perceptions of themselves as a spiritual/religious person and the importance of spiritual/religious beliefs in their how they conduct their life.

(8) Emotional Stability: Student self-perceptions of themselves as being calm and relaxed, their emotional stability, and how much they worry.

(9) General: Student self-perceptions of themselves as effective, capable individuals who have self-confidence and self-respect and are proud and satisfied with the way they are.

(10) Verbal: Student self-perceptions of their verbal skills, verbal reasoning ability and interest in verbal activities.

(11) Math: Student self-perceptions of their mathematical skills, mathematical reasoning ability and interest in mathematics.

(12) School: Student self-perceptions of their skills, ability and interest in school subjects in general.

(13) Problem Solving: Student self-perceptions of their ability to solve problems and think creatively and imaginatively.

Results.

Preliminary Analysis of the Psychometric Properties Of The SDQIII.

Self-concept research is notorious for the use of psychometrically weak instruments. This problem is particularly serious in the study of self-other agreement such as that reviewed by Shrauger and Schoeneman (1979). Without clearly establishing the psychometric properties of responses to a self-concept instrument for both self-responses and responses by others, there is no way to determine whether poor self-other agreement is due to the inability of significant others to accurately infer self-concepts, specific characteristics of the study, or the poor quality of the measurement instruments. From this perspective it is important to establish the psychometric properties of the measurement instruments for both self-responses and responses by others. The replicability of these characteristics across Australian and Canadian responses, in addition to its relevance to the evaluation of self-other agreement, is also a substantively important issue.

Reliability. Coefficient alpha estimates of reliability were determined for self-responses and other-responses in the Australian and Canadian studies. These four sets of reliability estimates were then compared with those reported in the SDQIII test manual (Marsh, 1992). All five sets of reliability estimates were similar and consistently high (Table 1). These results provide strong support for this aspect of the psychometric properties of the SDQIII responses and for the replicability of the results based on Canadian and Australian responses.

Insert Tables 1 and 2 About Here

Factor analysis. Separate factor analyses -- using iterated communality estimates, a Kaiser normalization, and an oblique rotation (SPSS, 1988) -- were conducted for self- and other-responses in the Canadian and Australian studies, and the results were compared to those based on the large normative archive (Table 2) presented in the test manual (Marsh, 1992). As described in the test

manual (Marsh, 1992) the 136 SDQIII items were used to form 68 item pairs (the first two items within each scale form one pair, the next two items within each scale a second item-pair, and so forth) that are the basis of factor analyses. This type of factor analysis was like that reported by Marsh, Barnes and Hocevar (1985), is recommended in the test manual, and is automatically performed by the scoring program that comes with the test manual. Target loadings are the factor loadings of the 68 item-pairs on the factor each is designed to measure, while all other factor loadings are nontarget loadings. In all five factor analyses (see Table 2), target loadings were high (medians of .67 to .72), nontarget loadings were consistently low (medians of .02 in all analyses), and correlations among the factors were low (medians of .07 to .09; also see Appendix I).

For purposes of the present investigation, self-responses and other-responses are each represented by 13 factor scores. The use of factor scores instead of simple summated scale scores was recommended in the original Marsh, Barnes and Hocevar (1985) study, by McCrae and Costa (1988) in their review of related research, and more generally in the SDQIII manual (Marsh, 1992). The factor scores were derived from the oblique factor analysis for the normative archive of responses by 2436 responses (Marsh, 1992) summarized in Table 2. Factor scores are a weighted total of responses to the standardized responses to each measured variable in which the weights correspond to factor score coefficients based on the regression approach (SPSS, 1988). Measured variables were standardized in relation to results from the normative archive and factor score coefficients were derived from the factor analysis of results from the normative archive. This is one of the procedures recommended in the test manual (Marsh, 1992) where all the relevant information needed to compute these scores is presented and is automatically implemented in the scoring program that comes with the test manual. Particularly for purposes of the present investigation, this procedure has the important advantage of providing factor scores that are comparable for self-responses and other-responses, and that are comparable in the Canadian and Australian samples. The use of normative results to construct factor scores is justifiable and generally preferable to factor scores based on results from one idiosyncratic sample (Marsh, 1992).

Although not used in subsequent analyses, sets of 13 factor scores were also derived from each of four separate analysis -- self-responses and other-responses in the two studies. Results from these separate factor analyses resulted in 52 (i.e., 13 SDQIII factors x 4 analyses) factor scores that were then correlated with those based on the normative archive computed by the scoring program. All 52 correlations were .97 or larger (median r=.99). This very high agreement between factor scores derived from different factor analyses provides further support for the replicability of the SDQIII factor structure across the different analyses.

These strong psychometric properties of SDQIII responses -- good internal consistency and a clearly defined, replicable factor structure -- are an important pre-requisite to subsequent comparisons between self-responses and other-responses that is frequently lacking in self-concept research and in studies of self-other agreement.

Insert Table 3 About Here

Self-Other (correlational) Agreement.

The major focus of this investigation is on self-other agreement based on the Canadian students and the comparison of these results with those based on the Australian study. For each study, a 26x26 correlation matrix (see Appendix I) represents correlations between the 13 SDQIII scales based on self-responses and the 13 SDQIII scales based on responses by significant others. Of particular relevance are the 13 correlations between self-responses and responses by others on matching SDQIII scales (i.e., the convergent validities) that are summarized in Table 3. In the Canadian study these vary from .400 to .769 (mean = .560) and are very similar to those from the Australian study that vary from .311 to .800 (mean = .568). By visual inspection, the pattern of convergent validities for the different SDQIII scales appears to be similar in the two studies. In order to provide a more objective index of this observation, the two sets of convergent validities were correlated with each other. The extremely large pattern correlation (r=.91) indicates that the relative size of correlations for different scales was very similar in the two studies. The four SDQIII scales on which self-other agreement was exceptionally high in the Australian study (rs of .74 to .80 in Table 2) also had the highest self-other agreement in the Canadian study, although self-other agreement on Parent Relations was smaller in the Canadian study. In summary, self-other agreement summarized here is very good -- much better than found in other self-concept research -- and remarkably similar in the Australian and Canadian studies.

MTMM matrices (see Appendix I) are traditionally evaluated with guidelines based on those proposed by the Campbell and Fiske (1959; see Marsh, 1988a). These are used to infer convergent validity, divergent validity, and method effects. Convergent validity is supported by statistically significant and substantial convergent validities (i.e., self-other agreement on matching SDQIII factors). Divergent validity is supported when all other correlations in the MTMM matrix are small, and smaller than the convergent validities. Undesirable method effects are inferred when correlations among different traits measured by the same method (e.g., correlations among self-ratings of the 13 SDQIII factors) are systematically higher than correlations among different traits assessed by different methods. Applying these guidelines to the MTMM matrix of correlations between self responses and responses by others in the Canadian and Australian studies (Appendix I) respectively indicated that:

1) All 13 convergent validities in each study were statistically significant and substantial (mean rs of .560 [Canadian] and .568 [Australian]).

2) For every one of the 312 possible comparisons between a convergent validity and another correlation in the same row or column of the square (heterotrait-heteromethod) block of coefficients, the validity coefficient (means of .560 and .568) was higher than the comparison coefficient (means of .047 [Canadian] and .039 [Australian] in each study (i.e., a total of 624 comparisons).

3) For the 312 possible comparisons between a convergent validity coefficient and other correlations in the same row or column of the two triangular (heterotrait-monomethod) blocks, the validity coefficient (means of .560 and .568) was higher than the comparison coefficient (means of .107 [Canadian] and .108 [Australian]) for 311 comparisons in each study.

4) For both studies, correlations among ratings by others (means of .112 [Canadian] and .130 [Australian]) and correlations among self-ratings (means of .101 [Canadian] and .0087 [Australian])

were slightly higher than heterotrait-heteromethod correlations (means of .047 and .039), suggesting a small method/halo effect.

These MTMM findings provide strikingly strong support for both the convergent and divergent validity of SDQIII responses. Once again, the results are remarkably similar for the Australian and Canadian studies.

Mean Differences

Although the major emphasis of this study is on self-other (correlational) agreement on the SDQIII factors, it is also relevant to examine mean differences. These were tested using a 2 (Country -- Canada vs. Australia) x 2 (rater -- self vs. other) MANOVA (SPSS, 1988) for the 13 SDQIII factors. There was a relatively large multivariate effect due to rater (($F(13, 1078) = 29.26$, $p < .01$, effect size = .26), a relatively smaller effect due to Country (($F(13, 1078) = 5.29$, $p < .01$, effect size = .06), and no significant rater x country interaction (($F(13, 1078) = 1.22$, $p = .26$)). Mean responses and supplemental analyses for each trait are presented in Table 4. Canadian responses are significantly higher for 1 trait, significantly lower for 1 trait, and did not differ significantly from Australian responses for the remaining 11 traits or the mean of the 13 SDQIII factors. Responses by others were significantly lower than self ratings for 1 trait, but were significantly higher on 6 traits and on the mean of the 13 SDQIII factors. (The one trait on which significant others gave lower ratings than the subjects was Parent Relations and many of the significant others were parents. Hence, even this one difference in which other ratings were higher may reflect a self-modesty effect on the part of parents.) The country x rater interaction was nonsignificant for the mean of the 13 traits and for all but 1 of the traits. Thus mean responses in the two studies are generally similar and in both studies and there is a tendency for ratings by others to be more favorable than self-ratings. Because the effect of rater is in the direction of higher responses by significant others in both studies, the results may suggest a modesty effect instead of a self-favorability bias. Because the modesty effect is similar in both studies (or slightly larger in the Canadian study), the results offer no support for the Marsh and Richards (1990) suggestion that such an effect may be idiosyncratic to responses by Australians.

Insert Table 4 About Here

Discussion and Implications.

The results summarized here have important theoretical and practical implications for self-concept research. The most obvious theoretical implication is for the study of ratings-by-others. However, interpreting the self-other agreement found here in terms of theory and previous research is difficult because of the ambiguous use of ratings-by-others and ambiguities in the theory underlying this research. The present findings apparently are consistent with the Shavelson et al. (1976) model, but Shavelson et al. intentionally de-emphasized the use of inferred self-concept responses and argued that they should not necessarily have any close correspondence with the preferred self-report measures of self-concept. Crandall (1973) suggested the pragmatic use of ratings-by-others as a means of validating self-concept measures, although the suggestions were apparently not based upon any specific theoretical position. The present findings are consistent with the symbolic interactionist perspective, but do not provide a comprehensive test of the theory (see Kinch, 1963; Marsh, Barnes &

Hocevar, 1985; Wells & Marwell, 1976). Results summarized here do, however, unambiguously demonstrate that significant others are able to accurately infer multiple self-concepts of a person who they know well, and this empirical relation is theoretically important.

MTMM studies based on the SDQI found significant self-other agreement between self-concept ratings by school children and self-concepts inferred by their teachers. Student-teacher agreement tended to be stronger in areas where teachers were able to make relevant observations (academic facets, physical ability, and peer relations), but was poorer for Parent Relations, Physical Appearance, and General self-concepts. Marsh and Craven (1991) found a similar pattern of relations self-concept responses and self-concepts inferred by teachers, mothers, and fathers. Surprisingly, the different significant others in this study did not differ substantially in their abilities to infer accurately those areas of self-concept most closely related to their respective roles -- Parent Relations for parents and academic factets for teachers. In the present investigation many of the significant others were parents, and self-other agreement on the Parents scales was very high. By this age, perhaps, parents and students are using more similar -- and, perhaps, more realistic -- criteria for evaluating self-concepts of Parent Relations. The lack of agreement on Physical Appearance found in the preadolescent research was unexpected. Perhaps, the standards used by teachers to infer Physical Appearance are different from those used by students, but even student-peer agreement on this factor was poor (Marsh, Smith & Barnes, 1984). This suggests that students may be using idiosyncratic standards in forming their own self-concepts of Physical Appearance and that these standards may not generalize to those that they employ in making ratings about one of their classmates. For responses by university students considered here, self-other agreement on Physical Appearance was much better, but still below the average for all scales. Here again it would seem that by this age respondents are using internal standards physical attractiveness that are more similar to those used by significant others. An important area for further research is to pursue this suggestion that as subject grow older the standards that subjects use for making internal self-evaluations become similar to those used by significant others and that this phenomena results in higher self-other agreement.

The juxtaposition between results of the present investigation and those of Marsh and Craven (1991) are particularly important. Unlike most previous research, self-ratings and ratings by others were based on responses to multi-item scales with strong psychometric properties in both studies. Both studies also reported stronger self-other agreement than earlier research based on subjects of a similar age. The substantially stronger self-other agreement found here with university-aged subjects compared with the preadolescents in the Marsh and Craven study supports the claim that self-other agreement is more substantial for older subjects.

Shavelson, et al. (1976) predicted that self-other agreement would be lower on general dimensions of self-concept near the apex of their hierarchy than dimensions closer to the base of their hierarchy that are more directly related to observable behavior. Based on their review of self-other agreement in personality research, McCrae and Costa (1988) also suggested that agreement would be higher on traits that are more observable. In terms of the Shavelson, et al. predictions, the two most general SDQIII scales are the General-Self and General-Academic scales. Averaged across the Australian and Canadian studies (Table 3) self-other agreement is lower for these two scales than any

other SDQIII scale. Because the General-self scale was added to the SDQI in its last revision, self-other correlations are available for only a few studies (see Table 14 in Marsh, 1988b). Based on these few studies, however, self-other agreement on the General-self scale is lower than for other SDQI scales. Because self-other studies using the SDQI are based on agreement between students and teachers, it is not surprising the that self-other agreement is higher on the general academic scale than some of the non-academic scales, but self-other agreement on the general academic scale is lower than for the either the Mathematics or the Verbal scale for the set of 8 MTMM studies presented by Marsh (1988). Thus, this prediction from the Shavelson model and the related proposal offered by McCrae and Costa (1988) is supported by results of previous SDQ research and the present investigation. The relatively poorer agreement on Physical Appearance, however, does not seem to fit this pattern in that physical attractiveness is clearly "observable." This suggests, perhaps, that the critical feature may be the concreteness of standards used to evaluate criteria relevant to different areas of self-concept and the extent to which these standards are shared by the subject and significant others.

The relatively poorer self-other agreement on the General-Self and, to a lesser extent, General-Academic scales supports theoretical predictions, but the practical implications of these results may be even more important. Historically, self-concept researchers relied almost solely on general or global scales and this practice is still prevalent. Marsh and Shavelson (1985), however, argued that self-concept cannot be adequately understood if its multidimensionality is ignored. In support of this contention, the mean correlation among the 13 SDQIII trait factors (see Appendix I) is only about .10, so that a single global dimension cannot adequately account for the specific scales. Consistent with this result, Marsh (1990a) argued that whereas there is support for a multifaceted hierarchical model of self-concept, the strength of the hierarchy is much weaker than originally assumed. Marsh and Shavelson argued if the role of self-concept is to better understand the complexity of self in different contexts, to predict a variety of behaviors, to provide outcome measures for diverse interventions, and to relate self-concept to other constructs, then multiple specific dimensions of self-concept are more useful than a single general scale. In support of these contentions Marsh (1990a) demonstrated that general self-concept tends to be less stable over time than other self facets, is less related to other external constructs than the specific dimensions most logically related to the construct, and is less sensitive to interventions designed to enhance self-concept than the specific dimensions that most closely match the intended outcomes of the intervention.

More recently, focusing specifically on academic self-concept, Marsh, Byrne, and Shavelson (1988; also see Marsh, 1990a; 1990b; 1992) made a similar distinction between general academic self-concept and more content specific dimensions such as Verbal and Math self-concept. Consistent with a large number of earlier studies, they found math and verbal self-concepts to be nearly uncorrelated in responses to each of three different instruments (the SDQIII and two other self-concept instruments). They also found support for the internal/external frame of reference model (Marsh, 1986) that was designed to explain why math and verbal self-concepts are not correlated. According to this model, students use an ipsative or compensatory process whereby accomplishments in mathematics are evaluated in relation to verbal skills, and vice versa. In support of this previous research, the two studies presented here found that Math and Verbal scales were negatively correlated

for self-responses (-.26 and -.11), for other-responses (-.06 and -.04). If Math and Verbal self-concepts are uncorrelated, or even negatively correlated, then they cannot be adequately explained by a single dimension of academic self-concept.

In summary, this research on self-other agreement contributes to a growing body of research calling into question the usefulness of the General-Self and General-Academic scales.

Self-other agreement found here is substantially higher than found in other self-concept research, and so it is informative to evaluate why. The apparent reasons are that: a) students were older (e.g., students knew themselves better or based their self-responses on more objective, observable criteria); b) both students and significant others made their responses on the same well developed instrument consisting of multi-item scales; c) self-other agreement was for specific characteristics rather than for broad, ambiguous characteristics or an overall self-concept; and d) the significant others knew the students better and in a wider range of contexts than the observers in most research. Marsh, Barnes and Hocevar (1985) demonstrated support for "b" by showing the self-other agreement and support for divergent validity were much weaker when based on single-item ratings than on multi-item scales. Results from both the Australian and Canadian studies support "c" in that self-other agreement is weaker for General Self and General Academic scales. Support for "d" comes from the comparison of the present results with the Marsh and Richards (1990) study in which the significant others did not know the students as well. Limited support for "a" comes from the comparison of the present results with those of the preadolescent (SDQI) studies, though there are so many differences that this comparison must be viewed cautiously. In related personality research, Funder (1987), Kenrick and Funder (1988), and particularly McCrae and Costa (1988) suggested that a similar set of characteristics would enhance self-other agreement on personality traits.

REFERENCES

Baldwin, J. M. (1897). Social and ethical interpretations in mental development: A study of social psychology. New York: McMillian.

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. Psychological Review, 81, 88-104.

Burns, R. B. (1979). The self-concept: Theory, measurement, development and behaviour. London: Longman.

Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. Review of Educational Research, 54, 427-456.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56,, 81-105.

Cheek, J. M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer-rating study. Journal of Personality and Social Psychology, 43, 1254-1269.

Cooley, C. H. (1902). Human nature and social order. New York: Scribner.

Conley, J. J. (1985). Longitudinal stability of personality traits A multitrait-multimethod-multioccasion analysis. Journal of Personality and Social Psychology, 49, 1266-1282.

Costa, P. M. & McCrae, R.R. (1988). Personality in adulthood: A six year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. Journal of Personality and Social Psychology, 54, 853-863.

Crandall, R. (1973). The measurement of self-esteem and related constructs. In J. Robinson & P. Shaver (Eds.), Measures of Social Psychological Attitudes (pp. 45-158). Ann Arbor, MI: Institute for Social Research.

Duval, S., & Wicklund, R. A. (1972). A theory of objective self-awareness. New York: Academic Press.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. Psychological Bulletin, 101, 75-90.

James, W. (1890/1963). The principles of psychology. New York: Holt, Rinehart & Winston.

Kenrick, D. T. & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. American Psychologist, 43, 23-34.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. Psychological Bulletin, 102, 390-402.

Kinch, J. W. (1963). A formalized theory of self-concept. American Journal of Sociology, 68, 481-486.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. American Educational Research Journal, 23, 129-149.

Marsh, H. W. (1988a). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), Educational research methodology, measurement and evaluation: An international handbook. Oxford, Pergamon Press.

Marsh, H. W. (1988b). Self Description Questionnaire: A Theoretical and empirical basis for the Measurement of multiple dimensions of preadolescent self-concept: A test manual and a research monograph. San Antonio, TX: The Psychological Corporation.

Marsh, H. W. (1990a). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. Educational Psychology Review, 2, 77-172.

Marsh, H. W. (1990b). The structure of academic self-concept: The Marsh/Shavelson model. Journal of Educational Psychology, 82, 623-636.

Marsh, H. W. (1992). The content specificity of relations between academic achievement and academic self-concept. Journal of Educational Psychology, 84., 35-42

Marsh, H. W. (1992). Self Description Questionnaire (SDQ) III: A Theoretical and empirical basis for the Measurement of multiple dimensions of late adolescent self-concept: A test manual and a research monograph. Faculty of Education, University of Western Sydney, Macarthur, New South Wales, Australia.

Marsh, H. W., Barnes, J., & Hocevar, D. (1985). Self-other agreement on multidimensional self-concept ratings: Factor analysis & multitrait-multimethod analysis. Journal of Personality and Social Psychology, 49, 1360-1377.

Marsh, H. W., Byrne, B. M., & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. Journal of Educational Psychology, 80, 366-380.

Marsh, H. W., & Craven, R. (1991). Self-other agreement on dimensions of preadolescent self-concept: Inferences by teachers, mothers, and fathers. Journal of Educational Psychology,.

Marsh, H. W. & Richards, G. (1990). Self-other agreement and self-other differences on multidimensional self-concept ratings. Australian Journal of Psychology, 42, 31-45..

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. Educational Psychologist, 20 107-125.

Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self-concepts: Relationships with inferred self-concepts and academic achievement. Australian Journal of Psychology, 36, 367-386.

Mead, G. H. (1934). Mind, self and society. Chicago: University of Chicago Press.

McCrae, R. R., & Costa, P. T. (1988). Different points of view: Self-reports and ratings in the assessment of personality. Paper presented at the XXIV International Congress of Psychology, Sydney, Australia.

Rosenberg, M. (1965). Society and the adolescent child. Princeton: Princeton University Press.

Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Validation of construct interpretations. Review of Educational Research, 46, 407-441.

Shrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the glass darkly. Psychological Bulletin, 86, 549-573.

SPSS (1988). SPSSx User's Guide. New York: McGraw-Hill.

Wells, L. E., & Marwell, G. (1976). Self-esteem: Its conceptualization and measurement. Beverly Hills, Calif.: Sage Publications.

Wylie, R. C. (1974). The self-concept. (Rev. ed., Vol. 1) Lincoln: University of Nebraska Press.

Wylie, R. C. (1979). The self-concept. (Vol. 2) Lincoln: University of Nebraska Press.

Table 1

Internal Consistency Estimates of Reliability Based on the SDQIII Normative Archive (N=2,436), Australian (N=151) Self-responses and Other-responses, and Canadian (N=941) Self-responses and Other-responses.

SDQIII Scales	SDQIII Archive	Australian Self	Australian Other	Canadian Self	Canadian Other
1) Physical Ability	.94	.96	.97	.96	.96
2) Appearance	.90	.86	.89	.90	.90
3) Opposite Sex Peers	.92	.90	.90	.91	.90
4) Same Sex Peers	.87	.86	.90	.85	.84
5) Parents	.89	.91	.93	.89	.89
6) Honesty	.74	.74	.81	.80	.82
7) Religion	.95	.95	.95	.94	.94
8) Emotional	.89	.91	.93	.89	.89
9) General	.93	.94	.93	.94	.92
10) Mathematics	.94	.95	.95	.95	.95
11) Verbal	.86	.84	.86	.86	.86
12) Academic	.92	.86	.88	.87	.89
13) Problem Solving	.84	.79	.82	.80	.81
Mean	.89	.88	.90	.89	.89

NOTE: Internal consistency refers to coefficient alpha estimates of reliability.

Table 2

Summary of Factor Analyses of SDQIII Responses For the Total Normative Archive (Norm Group; Marsh, 1992), Australian self-responses, Australian other-responses, Canadian self-responses, and Canadian other-responses.

	Norm Group (n=2436)	Aust Self (n=151)	Aust Others (n=151)	Canadian Self (n=941)	Canadian Others (n=941)
Target Loadings					
No. of Coefficients	68	68	68	68	68
Highest	.94	.92	.91	.93	.92
Lowest	.44	.23	.40	.40	.43
Median	.71	.67	.72	.69	.69
% > .30	100.0%	96.0%	100.0%	100.0%	100.0%
Non-Target Loadings					
No. of Coefficients	716	716	716	716	716
Highest	.25	.43	.36	.33	.30
Lowest	-.17	-.28	-.21	-.11	-.11
Median	.02	.02	.02	.02	.02
% > .30	0.0%	3.0%	0.1%	0.1%	0.0%
Factor Correlations					
No. of Coefficients	78	78	78	78	78
Highest	.36	.32	.33	.35	.36
Lowest	-.06	-.19	-.14	-.11	-.06
Median	.10	.07	.08	.09	.08
% > .30	5.1%	1.2%	1.2%	3.8%	2.6%

Note: As described in the test manual (Marsh, 1992) the 136 SDQIII items are used to form 68 item pairs (the first two items within each scale form one pair, the next two items within each scale a second item-pair, and so forth) that are the basis of factor analyses. Target loadings are the factor loadings of the 68 item-pairs on the factor each is designed to measure, while all other factor loadings are Nontarget loadings. Factor Correlations are the factor pattern correlations among the 13 oblique correlations identified in each analysis. The norm group includes the Australian self-responses but not any of the other responses considered here. Factor scores considered in subsequent analyses were derived from the factor analysis of the norm group as described in the test manual.

Table 3

Agreement Between Self-responses and Responses By Others in the Australian STudy (N=151) and the Canadian Study (N=941).

SDQIII Scales	Australian	Canadian
1) Physical Ability	.780	.754
2) Appearance	.461	.454
3) Opposite Sex Peers	.520	.587
4) Same Sex Peers	.457	.473
5) Parents	.759	.659
6) Honesty	.394	.437
7) Religion	.800	.745
8) Emotional	.624	.482
9) General	.403	.422
10) Mathematics	.741	.769
11) Verbal	.638	.614
12) Academic	.311	.440
13) Problem Solving	.490	.400
Mean	.568	.560

NOTE: Self-other agreement in both studies is based on factor scores (see Table 2). All correlations in both studies are statistically significant. When the set of 13 self-other agreement coefficients from one study were correlated with those in the other study, the correlation was r = .908

Table 4

Mean responses and the effects of Method (self vs. other), Country (Australia vs. Canada) and their interaction on responses to each of the 13 SDQIII scales (traits)

	Self-ratings		Other ratings		Significance Tests For:		
Score	Aust	Canada	Aust	Canada	Country	Method	Interaction
Physical	-.41	-.28	-.45	-.35	ns	ns	ns
Appearance	-.01	.11	.26	.32	ns	***[a]	ns
Opposite Sex	.31	.15	.31	.25	ns	ns	ns
Same Sex	.03	.15	.09	.29	*	*	ns
Parents	-.27	-.04	-.49	-.17	***[a]	***[a]	ns
Honesty	-.01	-.16	.02	.02	ns	*	ns
Religion	.16	-.11	.11	-.10	**[a]	ns	ns
Emotion	-.31	-.51	-.25	-.34	ns	**	ns
General	.12	.22	.15	.29	ns	ns	ns
Math	-.38	-.57	-.37	-.38	ns	**[a]	**
Verbal	.21	.08	.44	.35	ns	***[a]	ns
Academic	.24	.12	.21	.14	ns	ns	ns
Problem Solve	-.25	-.24	-.05	.03	ns	***[a]	ns
TOTAL	-.04	-.08	-.01	.03	ns	***[a]	ns

Note. Standard deviations corresponding to each of the mean responses is presented in appendix I.

[a] The tests of statistical significance reported here are based on an alpha = .05 level for each test and do not take into account the fact that multiple comparisons are made. These differences are statistically significant using the much more conservative Bonnferoni correction (i.e., the observed p-value x 13 < .05).

* p <.05, ** p < .01, *** p < .001.

Appendix I
Means, Standard Deviations, and Correlations Among the 26 Variables (13 SDQIII factors for self-responses and other-responses) from the Australian and the Canadian Studies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	Mean	SD
Canadian Study																												
Self-responses																												
PHYS	100	27	09	15	03	-11	-02	15	12	07	-05	00	12	75	19	11	14	00	-13	-04	14	05	04	-16	-08	02	-.28	1.30
APPR	27	100	35	10	05	01	02	14	42	-06	08	11	16	26	45	23	00	-03	-04	03	06	20	-06	03	03	06	.11	.93
OSEX	09	35	100	23	07	01	-09	18	31	-10	17	10	13	09	19	59	05	-02	-07	-04	01	17	-11	07	03	06	.15	.97
SSEX	15	10	23	100	18	08	02	24	24	-07	11	09	03	12	04	15	47	11	-01	05	15	16	-07	01	-02	-04	.15	.93
PRNT	03	05	07	18	100	19	20	18	16	-03	02	06	-05	04	03	-01	11	66	10	19	09	06	-01	-00	05	-05	-.04	1.01
HONS	-11	01	01	08	19	100	13	11	12	-06	18	22	04	-09	-07	-06	-03	10	44	11	-06	04	-06	16	16	03	-.16	.96
RELG	-02	02	-09	02	20	13	100	-01	09	-05	-04	05	01	-03	00	-07	05	19	10	74	02	06	-03	-03	04	-01	-.11	.94
EMOT	15	14	18	24	18	11	-01	100	37	04	11	12	12	11	05	13	17	12	00	03	48	26	02	02	04	03	-.51	.95
GEN	12	42	31	24	16	12	09	37	100	00	16	26	23	11	19	20	11	07	01	09	19	42	-03	06	08	11	.22	.97
MATH	07	-06	-10	-07	-03	-06	-05	04	00	100	-26	07	13	08	-06	-10	-05	01	-03	-04	07	-01	77	-16	12	15	-.57	1.19
VERB	-05	08	17	11	02	18	-04	11	16	-26	100	41	28	-07	02	13	04	-05	15	-06	02	15	-19	61	27	19	.08	.93
ACAD	00	11	10	09	06	22	05	12	26	07	41	100	22	-04	00	05	-02	00	15	03	-01	13	08	32	44	20	.12	.79
PROB	12	16	13	03	-05	04	01	12	23	13	28	22	100	04	07	08	-05	-09	-05	-05	07	15	07	10	14	40	-.24	.86
Significant Other Responses																												
PHYS	75	26	09	12	04	-09	-03	11	11	08	-07	-04	04	100	29	17	19	03	-06	-02	18	10	07	-10	01	10	-.36	1.31
APPR	19	45	19	04	03	-07	00	05	19	-06	02	00	07	29	100	30	14	07	00	06	11	26	-06	07	06	09	.32	.90
OSEX	11	33	59	15	-01	-06	-07	13	20	-10	13	05	08	17	30	100	24	-04	-09	-04	12	27	-08	15	08	14	.25	.89
SSEX	14	00	05	47	11	-03	05	17	11	-05	04	-02	-05	19	14	24	100	17	10	09	21	28	-02	13	08	07	.29	.89
PRNT	00	-03	-02	11	66	10	19	12	07	01	-05	00	-09	03	07	-04	17	100	23	23	17	13	06	03	09	-01	-.17	1.00
HONS	-13	-04	-07	-01	10	44	10	00	01	-03	15	15	-05	-06	00	-09	10	23	100	12	02	14	00	29	25	11	.02	1.03
RELG	-04	03	-04	05	19	11	74	03	09	-04	-06	03	-05	-02	06	-04	09	23	12	100	06	09	-03	-04	03	01	-.10	.90
EMOT	14	06	01	15	09	-06	02	48	19	07	02	-01	07	18	11	12	21	17	02	06	100	39	04	03	05	09	-.34	.97
GEN	05	20	17	16	06	04	06	26	42	-01	15	13	15	10	26	27	28	13	14	09	39	100	01	19	20	22	.29	.87
MATH	04	-06	-11	-07	-01	-06	-03	02	-03	77	-19	08	07	07	-06	-08	-02	06	00	-03	04	01	100	-06	32	25	-.38	1.14
VERB	-16	03	07	01	-00	16	-03	02	06	-16	61	32	10	-10	07	15	13	03	29	-04	03	19	-06	100	46	38	.35	.89
ACAD	-08	03	03	-02	05	16	04	04	08	12	27	44	14	01	06	08	08	09	25	03	05	20	22	46	100	35	.14	.81
PROB	02	06	06	-04	-05	03	-01	03	11	15	19	20	40	10	09	14	07	-01	11	01	09	22	25	38	35	100	.03	.84
Australian Study																												
Self-Responses																												
PHYS	100	12	-01	04	-08	00	08	16	01	-04	-22	-16	-02	78	07	10	-01	02	04	10	15	09	-12	-20	-01	01	-.41	1.28
APPR	12	100	11	12	-10	02	-08	-02	26	08	14	19	12	09	46	02	-00	-10	-00	-12	-05	13	-07	10	12	16	-.01	.86
OSEX	-01	11	100	27	19	18	-03	34	33	10	16	-02	02	-14	12	52	16	09	-02	-08	26	21	-00	02	-15	-06	.31	.90
SSEX	04	12	27	100	25	16	11	31	42	18	13	09	-01	-05	06	16	46	22	-07	03	18	21	06	-05	-02	-03	.03	.94
PRNT	-08	-10	19	25	100	09	13	16	12	12	10	-06	-22	06	11	10	23	76	08	06	25	19	14	15	-05	-11	-.27	1.14
HONS	00	02	18	16	09	100	13	11	13	01	12	14	-01	-12	12	-04	03	02	39	17	04	-02	-10	-08	-02	-00	-.01	.89
RELG	08	-08	-03	11	13	13	100	-05	02	04	-07	-03	-10	13	05	-01	09	19	29	80	07	10	09	-03	14	-10	.16	1.05
EMOT	16	-02	34	31	16	11	-05	100	43	06	08	-01	-03	17	06	36	27	17	05	-05	62	27	04	02	-16	13	-.31	1.03
GEN	01	26	33	42	12	13	02	43	100	12	07	22	12	00	21	19	19	12	02	-00	19	40	-05	-02	-03	08	.13	.90
MATH	-04	08	10	18	12	01	04	06	12	100	-11	09	07	01	04	00	06	12	-03	08	04	02	74	-09	32	09	-.38	1.10
VERB	-22	14	16	13	10	12	-07	08	07	-11	100	41	23	-20	-06	-09	02	-02	03	-10	08	05	-11	64	06	17	.21	.95
ACAD	-16	19	-02	09	-06	14	-03	-01	22	09	41	100	26	-10	-10	-19	-19	-14	03	00	-21	-01	08	24	31	11	.24	.73
PROB	-02	12	02	-01	-22	-01	-10	-03	12	07	23	26	100	-03	10	09	-16	-18	-02	02	-03	13	05	12	06	49	-.25	.92
Significant Other Responses																												
PHYS	78	09	-14	-05	06	-12	13	17	00	01	-20	-10	-03	100	19	14	02	13	12	16	17	15	01	-04	13	13	-.45	1.39
APPR	07	46	12	06	11	12	05	06	21	04	-06	-10	10	19	100	33	24	07	16	04	19	32	02	-02	-06	21	.26	.91
OSEX	10	02	52	16	10	-04	-01	36	19	00	-09	-19	09	14	33	100	38	09	-01	-04	46	40	01	-03	-06	20	.31	.88
SSEX	-01	-00	16	46	23	03	09	27	19	06	02	-19	-16	02	24	38	100	36	15	01	38	37	03	07	-03	01	.09	1.04
PRNT	02	-10	09	22	76	02	19	17	12	12	-02	-14	-18	13	07	09	36	100	19	14	28	28	17	08	04	-08	-.49	1.23
HONS	04	-00	-02	-07	08	39	29	05	02	-03	03	03	-02	12	16	-01	15	19	100	33	10	15	01	18	25	20	.02	1.07
RELG	10	-12	-08	03	06	17	80	-05	-00	08	-10	00	02	16	04	-04	01	14	33	100	03	07	09	-10	11	02	.11	1.09
EMOT	15	-05	26	18	25	04	07	62	19	04	08	-21	-03	17	19	46	38	28	10	03	100	40	-03	05	-05	14	-.25	1.07
GEN	09	13	21	21	19	-02	10	27	40	02	05	-01	13	15	32	40	37	28	15	07	40	100	-03	16	11	29	.15	.84
MATH	-12	-07	-00	06	14	-10	09	-04	-05	74	-11	08	05	01	02	01	03	17	01	09	-03	-03	100	-04	19	15	-.37	1.15
VERB	-20	10	02	-05	15	-08	-03	02	-02	-09	64	24	12	-04	-02	-03	07	08	18	-10	05	16	-04	100	34	36	.44	.89
ACAD	-01	12	-15	-02	-05	-02	14	-16	-03	22	06	31	06	13	-06	-06	-03	04	25	11	-05	11	19	24	100	11	.21	.78
PROB	01	16	-06	-03	-11	-00	-10	13	08	09	17	11	49	13	21	20	01	-08	20	02	14	29	15	36	11	100	-.10	.86

Note. Correlations are presented without decimal points. Each variable is a factor score that is scored in relation to responses to the SDQIII normative and standardized to have mean = 0 and SD =1. Thus, for example, Physical self-concept scores in both the Australian and Canadian university studies presented here are somewhat below average (mean < 0) and somewhat more variable (SD> 1) than responses in the normative archive.

BEST COPY AVAILABLE